DOCUMENT RESUME

ED 393 942 TM 024 994

AUTHOR Enright, Mary K.; Gitomer, Drew
TITLE Toward a Description of Successful Graduate Students. GRE Board Research Report No. 85-17R.
INSTITUTION Educational Testing Service, Princeton, N.J.
SPONS AGENCY Graduate Record Examinations Board, Princeton, N.J.
REPORT NO ETS-RR-89-9
PUB DATE Apr 89
NOTE 45p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Academic Achievement; Cognitive Processes; College Faculty; Communication Skills; Creativity; *Graduate Students; Graduate Study; Higher Education; *Interpersonal Competence; Personality Traits; Professional Development; *Simulation; *Student Characteristics; Student Motivation; *Success; Synthesis

ABSTRACT

A broad understanding of the skills and characteristics associated with successful performance in graduate school was developed through discussions with two groups of distinguished graduate faculty members, 15 in all. The first group consisted of psychologists with expertise in cognition and assessment. The second group was composed of faculty from other fields. These discussions had four major outcomes. The first was a characterization of the graduate education process as a form of apprenticeship that was suggestive of the kinds of skills and characteristics that contribute to success in many graduate programs. The second outcome was the identification of critical skills associated with scholarly and professional competence not currently measured by graduate admissions tests. A tentative list of the following seven competencies was developed: (1) communication; (2) creativity; (3) explanation; (4) motivation; (5) planning; (6) professionalism; and (7) synthesis. The last two outcomes concerned the assessment of these competencies through discipline-specific simulation testing and exercises that would allow students to display the identified competencies. An appendix lists the faculty consultants. (Contains 1 table and 42 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 393 942

GRE

GRADUATE RECORD EXAMINATIONS

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

H. I. BRAUN

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TOWARD A DESCRIPTION OF SUCCESSFUL GRADUATE STUDENTS

Mary K. Enright
and
Drew Gitomer

GRE Board Research Report No. 85-17R
ETS Research Report 89-9

April 1989

This report presents the findings of a research project funded by and carried out under the auspices of the Graduate Record Examinations Board.

EDUCATIONAL TESTING SERVICE, PRINCETON, NJ

TM024994

BEST COPY AVAILABLE

Toward a Description of Successful Graduate Students

Mary K. Enright

and

Drew Gitomer

GRE Board Report No. 85-17R

April 1989

Educational Testing Service, Princeton, N.J. 08541

Copyright © 1989 by Educational Testing Service. All rights reserved.

Acknowledgments

We thank all the distinguished faculty who served as consultants on this project. In addition, we are especially grateful to Lois Crooks, Walter Emmerich, Norman Frederiksen, Ann Jungeblut, and Sam Messick, who shared their ideas with us through thoughtful discussions and careful reviews of various versions of this report. Nevertheless, the views expressed in this report are those of the authors and are not necessarily endorsed by all the individuals who have contributed to the project.

Abstract

At present, the role of testing in education is being radically reevaluated. Advances in cognitive science suggest new approaches to assessment based on a better understanding of the nature of successful performance. Tests, that have been used primarily for prediction and selection are now viewed as having the potential to become more relevant to, and integrated with, the educational process. Such tests, however, require rethinking what should be measured and how it should be assessed.

In this report, a broad understanding of the skills and characteristics associated with successful performance in graduate school was developed through discussions with two groups of distinguished graduate faculty members. The first group consisted of eminent psychologists with expertise in cognition and assessment. The second group was composed of distinguished faculty from other fields, including the humanities, social sciences, and physical sciences.

These discussions had four major outcomes. The first was a characterization of the graduate education process as a form of apprenticeship that was suggestive of the kinds of skills and characteristics that contribute to success in many graduate programs. Based on this description, the second outcome was the identification of critical skills associated with scholarly and professional competence that are not currently measured by graduate admissions tests. A tentative list of seven competencies thought to be important for success in academic graduate programs and in subsequent professional roles was developed. The competencies include communication, creativity, explanation, motivation, planning, professionalism, and synthesis.

The last two outcomes concerned how these competencies might be assessed. One was a recommendation that discipline-specific simulation testing would be a useful tool that would permit the identification and concrete definition of skills associated with superior performance, contribute to the development of a theory of success in graduate school, and encourage the development of assessment techniques with well-established construct validity and educational relevance. Finally, potential exercises that would provide opportunities for students to display these competencies were proposed.

Table of Contents

Introduction 1

Part I: Factors Contributing to Success in Graduate School. 3
A Statement of The Problem 3
Setting the Context: The Nature of Graduate Education 4
The Nature of Expertise 6
Characteristics of Successful Graduate Students. 7
Important Competencies 9
An Alternative Approach to Assessment: Simulation Testing 13
Assessment Centers in Industry 14
Potential Exercises. 16
Implications of a Broader Conception of Graduate School Performance. 20
Summary. 22

Part II: Simulation Testing and Test Development. 22
Simulation Testing in Industry 23
Simulation Tests in Academia 26
Simulation Testing and the Development of Construct-Valid Tests. . 29
Concluding Comments. 32

References 33

Appendix 37

Introduction

Standardized testing evolved in a context in which the primary purpose of testing was prediction and selection. In the academic area, success was narrowly defined in terms of grades in school, and tests were evaluated in terms of their ability to predict grades. Dissatisfaction with this approach to testing has grown over the past two decades. This dissatisfaction stems from perceived limitations in current tests as well as an awareness of new possibilities for future tests.

Consequently, the role of testing in education is being radically reevaluated at present (Frederiksen, 1984; Glaser, 1981). Many factors have stimulated discussion of the purposes that assessment might serve in the future. On the one hand, these factors include misuses of standardized tests, inequalities in access to higher education, and concerns about a perceived decline in the quality of higher education. For example, the misuses of standardized tests include overrelying on the results of such tests in the selection and admissions process, allowing such tests to determine the curriculum, and using the tests to evaluate the educational effectiveness of programs and institutions. At the same time, concern about declines in the quality of education has increased demand for measures of educational achievement. The problem with such uses of many tests is one of a mismatch between the purposes for which the tests were originally designed and the purposes for which they are currently being used. A test designed as a valid predictor of an individual's classroom performance is not necessarily a valid measure of a program's or a teacher's effectiveness or of desired educational outcomes.

On the other hand, factors such as advances in cognitive science and the increasing use of computers in assessment suggest new approaches to assessment based on a better understanding of the characteristics of successful performance and more flexible methods of testing. Tests that have been used primarily for prediction and selection are now viewed as having the potential to more directly affect educational and instructional practices (Frederiksen, 1984; Glaser, 1981). Such tests, however, require changes in both the content of assessment instruments and the kinds of cognitive processes that are assessed. It is in the spirit of improving the quality of the information derived from testing that this research has been conducted.

The current project explores the relevance and construct validity of graduate admissions tests in the light of recent work in cognitive psychology and assessment. An attempt is made in this report to describe factors that may be important to, and predictive of, the development of cognitive skills requisite to graduate school success. A broadened concept of graduate school performance can contribute to the design of assessment instruments that not only improve prediction of future performance but are also relevant to the educational process. A second purpose of this project is to develop a research

BEST COPY AVAILABLE

approach that can serve as an agenda for the empirical evaluation of learning and performance in graduate school.

Our approach to this task was to discuss with two groups of faculty members what skills are thought to be critical to success in graduate school, how these skills are manifested in graduate performance in different fields, and how they might be assessed at various points during one's graduate career. These faculty members were noted scholars and mentors with direct experience in graduate teaching (see Appendix). The first group was drawn from the ranks of eminent psychologists with expertise in cognition and assessment; the second group consisted of distinguished scholars from other fields. Initial telephone conversations with individual psychologists provided the basis for a preliminary description of some characteristics of successful graduate students as well as identification of important issues. Subsequently, these psychologists met in a group to discuss the issues that had been identified and to suggest an elaborated framework for further exploration of student characteristics.

A major outcome of the first meeting was the suggestion that we consider how an assessment center approach with an emphasis on simulation testing might be used to explore skills important to success in graduate school. The potential utility of this approach became a central item on the agenda for our second meeting, at which we tried to identify skills and tasks important to success in graduate school that had generality across disciplines. While the idea of simulation testing was proposed by psychologists, it was well received by scholars from other fields, particularly the natural sciences, where such tests are sometimes used as course examinations or as part of cumulative or comprehensive exams.

This report is divided into two parts. In Part I, we formulate a description of successful graduate students. First, we briefly examine the problem of how the predictive and construct validity of current graduate school admissions tests are constrained by the limited range of behaviors that are currently assessed. A description of the process of graduate education sets the context for an attempt to identify important skills and tasks. Based on our discussions with faculty collaborators as well as a brief review of relevant research on expertise and on the characteristics of successful graduate students, we present a list of seven competencies important for success in academic graduate programs. We then discuss how the idea of simulation testing as used in assessment centers suggests appropriate assessment exercises. Potential simulation exercises that would afford opportunities for graduate students to display these competencies are then proposed. In the final section of Part 1, we consider the implications of this approach for theory, assessment, instruction, and guidance.

In Part II of the report we elaborate on the use of simulation testing in the test development process. Part II is divided into four sections. We selectively review how simulation tests have been used,

first, in industry, and, second, in academia, with a view toward illustrating both the limitations and the promise of this approach. Next, we discuss how simulation testing can be used to develop assessment instruments with good construct validity, paying particular attention to the work of Frederiksen (1986) in this area. We conclude with some comments about the potential outcomes of a research program based on simulation testing

Part I: Factors Contributing to Success in Graduate School

A Statement of the Problem

There is much that is yet to be understood about the factors that contribute to success in graduate school; the very nature of "good performance" in graduate school is ill-defined. The process of graduate education is complex, variable, and often unstructured. Although admission to graduate school is based predominately on measures of classroom performance or tests that predict classroom performance, evidence of potential professional distinction is typically exhibited and can be observed in situations outside the classroom.

The limitations of graduate school admissions tests in the face of the complexity of the graduate education process have long been recognized. Consequently, research to improve understanding of the nature of success in graduate school and of the characteristics of successful students has been conducted. Much of this research has been sponsored by the Graduate Record Examinations Board. For example, Hartnett and Willingham (1979) explored the criterion problem in the context of graduate education. In addition to traditional administrative criteria such as grades and progress toward the degree, they discussed the importance of other indicators of success, including evidence of professional accomplishment and specially developed measures. Other research has been directed toward describing the characteristics of successful students (Powers & Enright, 1987; Reilly, 1976; Tucker, 1985) or toward developing tests or inventories that measure a wider variety of student characteristics (Baird, 1979; Conrad, 1976; Crooks, Campbell, & Rock, 1979; Donlon, Reilly, & McKee, 1978; Frederiksen & Ward, 1978). Much of this work, however, has been concerned with specific aspects of performance and not designed to formulate a comprehensive picture of the qualities that contribute to success. Thus, the problem as we see it is the absence of an integrated overview of factors that contribute to success in graduate school and their relative importance, how such factors contribute to success, and how such factors are related to each other. Our goal is to work toward a more comprehensive description of what contributes to success in graduate school and to consider methods of substantiating this description.

BEST COPY AVAILABLE

Setting the Context: The Nature of Graduate Education

In meetings with faculty consultants, we discussed issues relevant to graduate education and student success. We also attempted to construct a preliminary outline of skills and corresponding tasks important in graduate education. In the following section, we describe the view of graduate education that emerged and the kinds of tasks and skills that appear to be important for success in graduate school.

Since the goals of graduate education include training of researchers, teachers, and practitioners, success in such different career paths may be determined by very different skills. However, in our discussions, the traditional view that the purpose of graduate education was to develop researchers who could produce new knowledge and communicate it to others was paramount. Inasmuch as most of our faculty consultants were involved in research-oriented programs, our discussions of factors important for success in graduate school emphasized research skills. We recognize that many, if not the majority of, graduate programs are more concerned with educating potential teachers or practitioners. However, in order to detail the skills important to success in such programs it would be necessary to hold discussions with appropriate faculty.

In research-oriented programs, graduate training can be viewed as a process of academic socialization. In particular, most graduate training is a form of apprenticeship. As students progress through their training, they are expected to move from apprenticeship to independence. From this perspective, success in graduate school is seen to be on a continuum with professional success, so that precocity in exhibiting behavior like that of a professional is considered to be a highly favorable sign. Hence, graduate school can be viewed as a work sample in which development as a student is equated with increasing approximation to professional behavior.

The primary way in which this socialization is accomplished is through modeling and participating in research. Many skills, such as writing, argumentation, and evaluation of research, are not formally taught and are acquired indirectly. Students also acquire practical professional knowledge this way. Modeling of these skills and opportunities for demonstration, practice, and critique often occur in lab meetings, tutorials, and seminars. Modeling also occurs in graduate courses where students learn the ways in which important field leaders think about problems and issues. However, successful modeling usually occurs in environments in which faculty are actively engaged in research activities. In part, this socialization process is facilitated by faculty selection of bright, motivated, and articulate students to work with them. Students who do not relate well to faculty and who lack sufficient social skills to benefit from modeling are clearly at a disadvantage in this process.

This view of graduate education suggests the kinds of tasks that are important for eventual professional success and that students should be exposed to during training. These include, for example, identifying significant problems for investigation; planning investigations of these problems; writing research proposals, papers, and reports; participating in collegial interactions and professional networks; and critiquing the ideas, proposals, and work of colleagues.

Traditionally, the first four of the above tasks are components of the master's thesis and the doctoral dissertation. These required tasks can be viewed as large-scale simulation tests that provide an opportunity for both training and assessment. Whether or not a student gains experience on other critical tasks is probably a function of the graduate program and the practices of the student's adviser, as well as the student's own interests, motivation, and personality.

In addition to content mastery and the skills of reasoning and writing that are traditionally assessed in graduate students, other characteristics that might contribute to success on these important tasks include interpersonal skills, oral communication skills, creativity, and motivation. Although many graduate faculty can agree on the importance of these other characteristics, overt assessment of such skills is not always considered an important component of graduate education. This may be due, in part, to a lack of reliable or standardized methods of assessing such skills or characteristics and, in part, to a lack of consensus that they constitute explicit goals of graduate education.

From the present perspective, it is desirable to develop simulation exercises that are of a smaller scope than the master's thesis or doctoral dissertation for several reasons. First, it is useful to know about a student's strengths and weaknesses prior to the huge investment of time and effort that accompanies a thesis project. Early diagnosis could result in interventions that help to avoid future problems that have high costs associated with them. Second, smaller tasks permit the analysis of limited numbers of separable skills, yielding more interpretable information. Third, given the often-subjective nature of graduate student assessment, it is possible that a more quantitative and statistically oriented instrument would be more valid. For example, Dawes (1971) found that statistically derived decisions about graduate school applicants were better predictors of graduate performance than were the clinical judgments of admission committees.

In the following sections we summarize the results of some relevant research on the nature of expertise as well as research on the characteristics of successful graduate students. This material and the comments of our faculty collaborators provide the basis for a preliminary taxonomy of competencies important for success in graduate school.

The Nature of Expertise

Graduate education is viewed here as a process of developing expertise in a given domain. The now substantial literature on expertise (e.g., Chi, Glaser, & Farr, 1988; Lesgold, 1984) provides a starting point for considering skills important to graduate success. Without reviewing the details of that research area, the following generalities emerge.

Experts have a great deal of domain-specific knowledge that is hierarchically organized in accord with the underlying structure of a domain (Chi, Feltovich, & Glaser, 1981). It is, of course, an understatement to say that graduate education almost always involves the mastery of a large, ever-increasing body of knowledge.

Central to any understanding of the knowledge base is the quality of mental models individuals have about a domain and about concepts or phenomena within that domain (cf. Gentner & Stevens, 1983). Models about a domain refer to the belief systems one has of a particular environment (Schoenfeld, 1985). These belief systems direct and shape one's cognitive activities. Thus, Schoenfeld gives an example of how an erroneous model of the mathematics world can manifest itself in weak problem solving. The belief that "formal mathematics has little or nothing to do with real thinking or problem solving" has the consequence that "in a problem that calls for discovery, formal mathematics will not be invoked" (Schoenfeld, 1985, p. 43). Because the demands and objectives of the graduate student researcher may vary significantly from previous scholarly experiences (predominantly classroom based), it is important that students develop useful models about the domain.

A second class of mental models refers to mental structures about concepts, phenomena, or principles central to a discipline. Models direct the use of information in problem solving and the acquisition of new information (Johnson-Laird, 1983). If an individual's model is not consistent with the structure of the domain, fallacious reasoning can occur in different contexts. Work by McCloskey (1983) in physics is one example of the naive theories people can have about a domain. Graduate students need to do more than acquire facts in their domain; they must develop models or structures that describe the interrelationships of concepts and systems within the discipline.

Current instruments (e.g., the GRE Advanced tests) can sample parts of this declarative knowledge base. However, standardized instruments are not presently in use that assess the rich organizational aspects of that knowledge, including the features of individuals' mental models. Inferring relations between concepts, building new knowledge, and accessing knowledge in problem-solving contexts are all a function of having well-organized knowledge and well-developed mental models (Johnson-Laird, 1983). Thus, one dimension of skill to be considered in a description of graduate

success should focus on the nature and quality of the organization of the knowledge base.

A second general feature of the expert is mastery of numerous procedures that enable one to navigate the domain. Not only do experts know these procedures, but they are able to select the most appropriate ones for a given situation. Thus, for any graduate domain, a set of critical procedures for accessing and combining or reconfiguring knowledge must be identified, along with the conditions or heuristic principles that suggest their use.

A third aspect of expertise, and perhaps the most important, has to do with those higher-level processes that organize and direct the processing of both declarative and procedural knowledge in problem-solving contexts. These processes control the use of appropriate knowledge, the execution of procedures, and the building and satisfying of goal structures in developing a problem representation. Problem representation has been shown to be critical in the development of expertise in domains as disparate as physics (Larkin, McDermott, Simon, & Simon, 1980) and the social sciences (Voss, Greene, Post, & Penner, 1983).

Characteristics of Successful Graduate Students

Few researchers have systematically explored skills important to success in graduate school. Much of the research that has been conducted in this area and is described below has been sponsored by the Graduate Record Examinations Board. Reilly (1976), using Flanagan's (1954) critical incident technique, queried a large number of faculty members from chemistry, English, and psychology departments. Faculty were asked to rate an average, below-average, and above-average student in terms of a checklist of behaviors that could describe student performance. In Reilly's analysis, the factor accounting for the most variance in each of the three disciplines was construed as "independence and initiative." Original research, independent execution, and self-directed learning were behaviors that loaded heavily on this factor. Frederiksen and Ward (1978) developed tests of scientific thinking to assess creative problem solving. One important finding was that these measures of creativity were better predictors of professional involvement than were more traditional ability and achievement test scores.

A long-range, systematic research program on the prediction of career progress for graduate students in management, sponsored by the Graduate Management Admission Council, probably represents the most comprehensive study of factors that contribute to success in graduate school and thereafter. In a series of studies, alternative criteria of success such as faculty ratings of graduate students (Hilton, Kendall, & Sprecher, 1970), an in-basket simulation test (Crooks, 1971), and measures of career progress after graduate school (Crooks & Campbell, 1974) were developed. The faculty ratings included scales such as Communication Skills; Critical Awareness; Initiative,

BEST COPY AVAILABLE

Persistence and Drive; Perspective and Breadth of Knowledge; Planning; and Problem Analysis. In the final study in this series, Crooks, Campbell, and Rock (1979) investigated the relationship of predictor variables such as undergraduate and graduate grades, graduate admission test scores, biographical and background information, faculty ratings, and personality measures to indicators of career progress obtained seven years after the completion of graduate school. Indicators of career progress included measures of job mobility, salary and salary progress, level and type of job responsibility, and job satisfaction. One important finding was that measures of career progress were better predicted by variables represented in the faculty rating scales and by other measures of motivation, interests, and personality than by academic ability and achievement measures. A second important finding was that predictor-criterion relationships differed for subgroups following different career paths. For example, faculty ratings of graduate students' perspective and breadth of knowledge was associated with career progress positively for those in staff or advisory positions and negatively for those in specialist positions (e.g., technical, research).

Baird (Baird, 1979, 1985; Baird & Knapp, 1981) developed an inventory to assess the prior accomplishments of graduate school applicants in a systematic way. The inventory consisted of a check-list of a wide variety of activities, such as writing and/or publishing fiction or scientific articles, entering literary or artistic contests, building mechanical or electronic devices, making clothes or handicrafts, fund raising, and holding offices in organizations. In addition there were some open-ended questions about prior achievements. Baird and Knapp (1981) administered the inventory to a sample of recently admitted graduate students in biology, English, and psychology and collected follow-up data about graduate grades and accomplishments at the end of the first year of study. Four clusters of accomplishments were abstracted from the responses to the inventory. These included clusters of literary and expressive activities, scientific and technical activities, artistic activities, and social-service organizational activities. Overall, pre-graduate school accomplishment predicted graduate school accomplishments in an appropriate manner. However, correlations of graduate school grades with both pre-graduate accomplishments and graduate school accomplishments were negligible, as were correlations between undergraduate grades and pre-graduate accomplishments. Thus, this inventory appeared to provide information about relevant skills that, although unrelated to grades, were nonetheless predictive of certain kinds of valued accomplishments in graduate school.

A number of recent studies have attempted to define reasoning skills more precisely, specifically as they apply to graduate education. Tucker (1985) compiled a list of general reasoning processes that could be important for graduate education and had expert philosophers and cognitive psychologists rank them in order of importance for graduate study. The reasoning processes described were suggested by the work of cognitive psychologists as well as

philosophers. The processes ranked as most important included Formulating Alternatives, Noticing (significant aspects), and Finding Vulnerable Parts (of a theory or plan).

Powers and Enright (1987) had graduate faculty in six disciplines judge the importance of various analytical reasoning skills in graduate performance. These included skills judged as important to all disciplines (e.g., reasoning in situations in which all the needed information is not known) as well as skills thought to be critical in only subsets of disciplines. For example, chemistry faculty placed high value on being able to generate hypotheses and to draw sound inferences from observations, while English faculty placed more importance on skills central to argumentation.

The above studies had different objectives and varied greatly in their approaches to documenting skills or characteristics associated with success in graduate school. There is important convergence, however, in work as disparate as that of Frederiksen and his associates and of Baird and his colleagues, namely, that it is possible to assess some skills that are distinct from those tapped by traditional ability measures. Furthermore, when the criteria are something other than grades (e.g., professional behaviors or accomplishments), these skills are more predictive than are traditional ability tests.

However, for our purposes, a limitation that applies to much of this work is that it is not immediately obvious what an assessment task would look like, given the abstractness of the concepts posited as important. While converging evidence exists that formulating alternative hypotheses is a critical graduate student skill, unless the construct is instantiated in specific tasks, it could be argued that formulating alternative hypotheses is equally important for the second grader as it is for the second year-graduate student. By instantiating the construct within a specific context, one can differentiate more precisely the skills needed by individuals facing varying demands, even though the same nomenclature may be used at different levels.

A second problem is that much of this research has, at least implicitly, treated process constructs as being essentially knowledge-independent. Yet much recent research has demonstrated that sophisticated information processing in a domain is facilitated by having highly developed domain knowledge (Chi, et al., 1988; Johnson-Laird, 1983).

Important Competencies

On the basis of the above material and our discussions with graduate faculty, we distilled the following list of seven general competencies that seemed to differentiate the more successful graduate students from the less successful. We turn now to brief descriptions of these competencies, and examples or occasions in graduate education

when these skills are manifest. The competencies in the list are presented in alphabetical order, not order of importance. However, the competencies that seemed most critical to many of our faculty consultants were creativity and motivation.

Communication. The essence of communication is the ability to share one's ideas, knowledge, and insights with others. The goal of communication forces an individual to organize and apply numerous skills. One must be able to reason from different viewpoints, follow appropriate communicative protocols, and tailor the communication to the audience. In addition, one must be able to comprehend and respond to the communications of others. Communication can be formal or informal, written or oral, spontaneous or planned.

However, communication skills may differ in form from one domain to another. The interpretive writing of the sociologist has a wholly different communicative style from the step-by-step proof of the mathematician or the dialogue of a playwright. Even within a domain, one does not communicate with peers in the same way as one does with a lay audience. The accomplished scholar must understand his audience and attempt to fulfill the implicit contract between the communicator and the receiver (Grice, 1975). Thus, domain-specific techniques and styles of communication must be mastered as well as generic communication skills (Bartholomae, 1986).

Communication is not unidirectional. Students need to be able to solicit and use criticism in the course of developing ideas. They need to learn how to benefit from having their ideas confronted so as to move on to better formulations on the basis of criticism. Thus, it may be useful to assess students on occasions in which they receive feedback from others and to determine the effects of such feedback on student performance.

Creativity. Traditionally, creativity has been assessed in terms of the ability to produce an unusual number of ideas or to generate novel ideas, but it has many other connotations. These include curiosity and intellectual playfulness or rebelliousness. Intellectual playfulness or rebelliousness refers to the ability to recognize that facts, concepts, and theories can be subjected to criticism, revision, modification, and reinterpretation. Thus, creativity also requires a domain model that permits this sort of intellectual playfulness.

Creativity is possible in connection with almost any problem-solving task in graduate school. Whether one is planning research, interpreting data, writing a paper, or giving a talk, one can proceed in relatively creative or in more prosaic ways. Thus, creativity appears to be an important aspect of student functioning that cuts across tasks and assessment dimensions.

Explanation. Explanation is the giving of a reason or cause for some phenomenon or finding. This class of skills is critical in the

interpretation and analysis of any research, either one's own or someone else's. Explanation can involve divergent production of alternative hypotheses, evaluation of competing hypotheses, or development of an argument that supports an explanation that needs to be communicated. Explanation can also be important in the development of self-knowledge that can influence student decision making.

Explanation obviously requires developed reasoning skills. Among the more traditional ability and achievement dimensions, various aspects of reasoning appear to be particularly important. Some of these aspects include analogical thinking in a broad sense (for example, being able to see how the research paradigms used to explore one topic might be applied to the investigation of a very different area), the ability to develop a logical chain of argument, the ability to use the appropriate argument structure or logic structure of one's field, and the ability to defend one's ideas. Evidence of explanatory competence in graduate school emerges in students' writing, in research design, in oral presentations both in class and at colloquia, and in informal interactions.

Motivation. Successful students are characterized by commitment, involvement, and interest in their work. Motivation may be demonstrated by persistence in working on problems, by enthusiasm and excitement about work, by pursuing problems or assignments beyond the minimum required, and by attending nonrequired colloquia and professional meetings.

The problem of evaluating applicants' motivation often centers around assessing their interests, which should be of an appropriate degree of specificity (neither too broad nor too narrow) and consistent with what the department has to offer. A history of productivity or independent achievement in any area may be evidence of independent, self-activated scholarship.

Some of our faculty consultants expressed the strong view that the differences between successful and unsuccessful students are motivational rather than cognitive. According to this view, the reason many students leave graduate school is that they understand neither the type and amount of work required nor the degree of commitment necessary to succeed. (This view was also fairly widespread among the graduate faculty whom Hartnett and Willingham [1979] interviewed nearly a decade ago.) In contrast, motivation was viewed by other faculty consultants as an important but not the determining factor in graduate success. Motivation is important, for example, in increasing student involvement in professional activities and thereby providing more opportunities to observe, acquire, or practice skills critical to success. The issue of how motivational variables are systematically related to cognitive variables needs further elaboration.

Planning. Planning refers to the development of a procedure to reach some goal. Planning is involved in such diverse activities as

BEST COPY AVAILABLE

designing a research experiment, organizing a paper or presentation, devising a solution to a mathematical problem, and making career decisions. Planning often includes identifying problems or formulating topics in a manageable way, devising strategies to answer questions, deciding what kinds of evidence are needed to resolve a question, and anticipating likely problems or criticisms. Planning is not just a preparatory activity that occurs at the beginning of an undertaking. Rather, it is an ongoing, interactive process. Planning must be flexible and responsive to new data, ideas, and perceptions.

Professionalism. Professionalism refers to skills in successfully accommodating to the social conditions of a particular field. Included here are social skills, knowledge of the communication channels and the power structure of the field, and practical knowledge. For graduate students, social skills are reflected in the ability to relate to faculty, to reach out to them, and to treat them as colleagues and equals. Knowledge of the communication channels and power structure of the field includes knowing how to find relevant research and the most recent unpublished reports, which conferences to attend, what professional societies to join, and which journals to read regularly. Finally, practical knowledge involves understanding what is being rewarded by the environment and adapting to it (Sternberg, 1985). In short, the student needs to develop a model of the profession that is consistent with the mental models possessed by successful professionals. Many indicators of a student's professional "savvy" emerge in graduate school. Their professional social skills are evident in their participation in informal seminars, in informal discussions with faculty and other students, and in interactions with visiting speakers. Joining professional societies, reading the appropriate journals, and attending and participating in conferences are also good indicators.

Synthesis. Every domain has an extensive knowledge base that needs to be mastered. Mastery is not simply the accretion of discrete facts, but the organization of this information into complex knowledge structures. For our purposes, synthesis refers to those skills that facilitate the development of expert domain knowledge structures.

Well-synthesized knowledge signifies a firm grasp of the subject area, an understanding of major theories, and a mastery of the content and skills of a field. However, synthesis also implies the capacity to function with a degree of independence, to be able to manipulate knowledge creatively, and to apply available skills under appropriate conditions. Learning has to involve more than just a rote accumulation of facts and research skills. As knowledge is acquired, students have to be able to organize it and then reorganize it to make inferences. Thus, synthesis may be evident in a creative restructuring of knowledge in an area, in the identification of new problems to be investigated, or in the development of new approaches to old problems.

Exactly what synthetic approaches are valued may vary across domains as well as among individuals within a domain. Opinions differ with respect to the relative value of depth versus breadth of knowledge. Some emphasize the importance of breadth for its contribution to serendipity, while others prefer depth and emphasize the importance of students becoming experts in their specific areas of interest.

Summary. We have presented a provisional list of seven competencies that appear to be important for success in graduate school. Although there skills are often evaluated in a relatively unsystematic manner, some evidence for many of them may be obtained from students' undergraduate records, from letters of recommendation, from faculty evaluations, and from students' own products such as papers or reports. However, all of these sources of information have certain drawbacks. For example, letters of recommendation are typically unstandardized and of unknown reliability. Faculty evaluations are likely to be influenced by how well a faculty member knows a student. In considering student reports and papers, it is often difficult to separate students' contributions from those of their advisers. Given these existing sources of information about student performance, we turned our attention to the problem of how to best assess the very complex competencies described above in a more standardized and systematic manner.

An Alternative Approach to Assessment: Simulation Testing

In our discussions with graduate faculty, we tried to develop recommendations about the kinds of assessment instruments that would permit students to demonstrate the competencies described above. Assumptions that influenced the direction of our discussions included the following:

- o the purposes of assessment should not be limited to prediction but should also include diagnosis, instruction, and guidance;
- o it would be difficult, if not impossible, to improve graduate assessment if we limited ourselves to thinking in terms of timed tests in a multiple-choice format;
- o new methods of assessment should be embedded in discipline-specific content matter.

A major outcome of our discussions was the recommendation that an "assessment center" approach might suggest appropriate kinds of exercises and methods of assessment.

The idea of simulation testing emerged as a central theme in our discussions for a number of reasons. Given that complex performance is best observed under conditions of realistic complexity, a reasonable approach to formulating a description of important

cognitive processes in graduate study would be to use problem simulations or work samples as a data collection device. In simulation testing the complex, ill-structured, and dynamic nature of real-life problem solving can be maintained while the test situation can be standardized enough so the data can be meaningfully interpreted (Frederiksen, 1986). Simulation tests, standardized and abbreviated versions of tasks critical for success in a field, can provide a work sample that can be analyzed for indicators of successful performance. The development of such tasks would facilitate the definition and measurement of skills and characteristics associated with success in graduate school in very concrete terms. In addition, simulation testing provides an organizational frame for thinking about the nature of graduate education, the characteristics associated with success in graduate school, and important tasks to be accomplished in graduate school. Finally, a research approach using simulation testing as a part of the test development process has been articulated by Frederiksen (1986).

Since simulation testing has been highly developed in industry in assessment centers where the main function is to help select and train managers, we briefly describe their operation below.

Assessment Centers in Industry

Assessment centers are not necessarily physical locations but instead are a set of procedures used to evaluate individuals for a variety of purposes, including training, development, prediction, and promotion (Thornton & Byham, 1982). The assessment center approach is characterized by clear identification of the dimensions to be assessed and the use of multiple assessment techniques. These elements of the assessment center approach are described briefly below. A systematic job analysis is usually conducted to determine specific activities that are important aspects of a job and to identify the characteristics necessary to carry out these activities effectively.

Behavioral Dimensions. In the assessment center approach, competence is defined as a set of observable behaviors grouped into a number of categories or dimensions. These dimensions are not necessarily traits in that there is no assumption that the identified behavioral consistencies represent stable, underlying psychological constructs.

The process of formulating dimensions is a somewhat subjective enterprise. In the context of a given assessment, dimensions are defined in detail and examples given of how they might be evidenced in a specific job. Planning and organizing in a manager, for example, might be generally defined as "ability to efficiently establish an appropriate course of action for self and/or others to accomplish a specific goal; make proper assignments of personnel and appropriate use of resources." (Thornton & Byham, 1982, p. 140). A specific example of good planning in a particular job is a supervisor who checks each morning to see if his whole crew is in and then makes

appropriate changes in work assignments to compensate for absences and to ensure that critical tasks are accomplished.

Assessment Techniques. The assessment techniques are multi-faceted, typically including paper-and-pencil tests, structured background interviews, and job simulation exercises. Structured background interviews are conducted by trained assessors and designed to elicit specific evidence relevant to the behavioral dimensions being evaluated. Simulation tests, in general, are complex performance tests carried out in real life or lifelike settings and, therefore, vary greatly with the assessment context. Some examples of exercises that have been used in various settings, particularly for manager positions, include the following.

1. Management games and leaderless discussion groups are exercises in which individuals have to work together in groups to solve a problem. This kind of task is often used in industry and provides opportunities for assessors to rate participants on such characteristics as leadership, oral communication, analysis, and planning skills.

2. In-basket exercises are simulations of the administrative aspects of many management jobs. Typically the participant is given a variety of documents such as letters, memos, and phone messages. The task is to respond to this material either by making notes of what should be done or by writing responses. At the end of the exercise, the participant may be asked to explain the reasons for taking certain actions. Dimensions that are assessed on this task might include ability to set priorities, efficiency, planning, and decision making.

3. Case study analyses are tasks in which the participant is provided with data about a situation and asked to recommend a course of action. The content of the task is usually modeled around important aspects of a particular job and might involve financial analysis, marketing strategies, or personnel problems. The participant might be asked to propose a course of action or discuss possible solutions in a group with other candidates. Skills that can be observed in this type of exercise include analysis, judgment, communication, and creativity.

4. Interview simulations involve the participant in the interviewing of simulated patients, clients, or subordinates. This type of exercise has been used in medical education as well as industry and provides opportunities for assessors to observe a participant's interpersonal skills, analytic ability, and communication skills.

The elements of the assessment center approach described above helped us to focus our discussions with graduate faculty. With them, we considered what simulation tasks relevant to graduate education might be. The results of these discussions are presented in the following sections.

BEST COPY AVAILABLE

Potential Exercises

The following are thought of as generic assessment exercises that may have utility across disciplines. For each discipline, however, the exercises will need to be adapted so they are consistent with the actual demands of the discipline. Each exercise is designed to provide the opportunity to evaluate student performance in terms of at least one of the identified competencies. They also share the feature that successful completion will usually require a considerable amount of time. These exercises are not intended to duplicate the writing of a thesis; they are intended to sample, efficiently and representatively, many of the same skills required in that sort of prolonged effort. In particular, there is no requirement that the exercise be completed within time limits typically associated with standardized testing (e.g., three hours). Rather, these exercises are to be treated as serious problem-solving activities or extended projects.

Table 1 illustrates the hypothesized relationships between the proposed exercises and the competencies identified. As can be seen, the relationship between exercises and competencies is not one-to-one. Ideally, evidence for each competency should be sought on a minimum of two exercises, while each exercise should provide an opportunity to observe a number of competencies. One point that should be kept in mind when examining this matrix is the tentative nature of both the competencies and the potential exercises described below. They are intended as illustrations that need to be evaluated and refined through further discussion and research.

Another constraint is that not all of these competencies or exercises are equally important in all disciplines. Disciplines vary greatly in the degree to which there is consensus about standards of acceptability or adequacy (be they standards of truth, beauty, or worth). For example, in mathematics there is a great deal less controversy about whether the proposed resolution of a problem is acceptable than there is in history or literature. Thus, certain types of critical and evaluative skills are more important in many of the humanities than in mathematics. On the other hand, identifying important problems for study is a skill that is very important in many diverse disciplines.

Exercise 1 - Structured Background Interview. A structured background interview provides an opportunity for the interviewer to gather information about individuals' interests, previous academic experiences and achievements, and sources of enjoyment and satisfaction. Such interviews are structured or standardized to elicit behavioral evidence from each individual relevant to specific behavioral dimensions. In the context of graduate school, students might be interviewed about their research interests, previous academic and nonacademic accomplishments, and the way they spend their free time. A structured interview is not, of course, a simulation exercise, but it provides an opportunity to obtain information about competencies such as oral communication skills, professional

Table 1

Competencies Expected to Be Evident

on Various Exercises

Competencies	Exercises				
	Background Interview	Complete Report	Plan Research	Critique Report	Peer-Group Discussion
Planning		X	X		X
Communication					
Oral	X				X
Written		X	X	X	
Explanation		X	X	X	X
Synthesis		X	X	X	X
Professionalism	X			X	X
Motivation	X				X
Creativity		X	X	X	X

involvement, and motivation that is difficult to obtain through other methods.

This task is viewed as most appropriate for the applicant or entering student. Often, the expectations of the beginning graduate student and the reality of graduate school and the profession are inconsistent. This lack of congruity often has deleterious effects on student motivation. Also, students may not apportion their resources in optimal fashion (Sternberg, 1985). Therefore, assessment information from this task can be used in two ways. First, students who do not have a realistic view of graduate school can decide on the basis of feedback whether or not such an environment suits their needs. Similarly, student search committees can use the information in an analogous fashion. Second, students and faculty can modify their learning and teaching approaches, respectively, with the hope of reducing the inconsistency between the reality and the student view.

Exercise 2 - Complete a Report. Candidates would be presented with an unfinished report about some recent work in their field for which they have to write an interpretation and discussion section. This would provide an opportunity to evaluate students' abilities to plan, to communicate in writing, to synthesize information, and to devise explanations for findings. In addition, creative handling of each of these problem aspects can also be evaluated.

Such an exercise has relevance throughout the graduate and professional career. A researcher in almost any discipline must be able to synthesize and interpret information, as well as offer and communicate explanations for such a set of findings. One can also assess the extent to which this is accomplished in a creative manner.

In completing a report, as in other exercises, a proper level of required domain knowledge must be achieved. Although novel information may be presented in the exercise, the examinee should be able to use domain information that is generally accepted as being "core" to the field to complete the task. However, the knowledge required should not be so specific that a bias exists favoring subsets of individuals who have had special exposure to relevant material.

Exercise 3 - Plan Further Research. In conjunction with completing a report, students could be asked to outline further research that could be done either to clarify the results of the report or to address new issues suggested or overlooked in the report. Students would need to identify the issues or questions to be pursued, suggest possible methods for investigating these issues, recognize the kinds of evidence that would be relevant, anticipate possible outcomes, and suggest likely explanations for alternative outcomes. Evidence about students' planning skills, creativity, and synthetic skills would be obtained from this exercise. Additionally, important methodological skills of a discipline could also be evaluated.

Constraints and considerations bearing on this exercise are similar to those for the previous task. While important to all research participants in a field, the activities required in this exercise must not depend on an extensive background in any part of the field that is familiar to only a few. Rather, the exercise should use as a model published articles that appear in journals having interest for a broad range of individuals across the discipline, not just specialists in a subfield.

Exercise 4 - Critique a Paper or Report. Students might be asked to write a review of a paper for a colleague, offering suggestions about what needs to be done to make it publishable and about likely publication outlets. Alternatively, examinees might be asked to review an article submitted to a journal or to referee conflicting reviews of a report.

This type of task also has relevance throughout the research career. The competent reviewer must be able to analyze a paper at multiple levels--from theoretical, methodological, and communicative perspectives at least. The review itself also provides a forum for displaying communication skills. The manner or style of these communications and suggestions for improvement provide information vis-a-vis interpersonal skills and professional awareness or adaptiveness.

Exercise 5 - Peer-Group Discussion. One of the above tasks, such as the research plan or the critique, might serve as the stimulus for a discussion by a group of students. Here, students' oral communication skills, as well as their creativity, their involvement in problem solving situations (motivation), and their ability to apply their domain knowledge to a problem (synthesis), might be evaluated. Group discussions, however, present a particularly difficult problem in terms of standardization. Different groups would vary in the types of personalities present. For example, the presence of a very talkative or dominating individual would drastically curtail the opportunities for others to participate. One way in which this kind of variation could be reduced would be to present a videotaped group discussion. A candidate's oral responses would be recorded either at specified points in the discussion or as the candidate thought appropriate.

Summary. The competencies and potential exercises we have described incorporate many of the ideas that emerged in our discussions with graduate faculty. No doubt other graduate faculty would feel that competencies we have listed are ones that they already evaluate in their students, albeit in an informal manner. They might also agree that the exercises we have suggested are similar to situations in which they have observed critical aspects of performance. However, they are likely to question what benefits might emerge from a program of research based on such an approach. In the following section we elaborate on the implications of a clearer definition of factors that contribute to success in graduate school.

Implications of a Broader Conception of Graduate School Performance

Assessing graduate student performance through the approach described above is based on the premise that the development of new methods should be driven by a theory of the processes that contribute to success in graduate school. Therefore, theory development is deemed as important as the development of new test items. However, in addition to theory development, this approach would have implications for many other aspects of graduate education, including prediction, assessment, instruction, guidance, and communication. The relevance of such an approach for theory development as well as for other educational functions is described below.

Theory. Central to this report is the problem of how to integrate a description of successful graduate students into a comprehensive theoretical framework that would facilitate the evaluation of graduate student achievement as well as the prediction of success in graduate school. Some of the questions that such a theory of successful graduate student performance might address include the following: Are certain skills prerequisite for the development of other skills? Can certain skill strengths compensate for weaknesses? Are certain skill levels sufficient? Necessary? Are these skills general or domain specific? To what extent can these various skills be trained? Is it desirable to do so? What, if any, changes in graduate education might be suggested? How early, and how reliably, do different skills manifest themselves in education (undergraduate or graduate)? Which skills need to be developed before entering graduate school, and which are developed as a part of the graduate education process?

A related topic concerns how students can optimally capitalize on their available skills. Many students and professionals will perform certain tasks better than others. Some may be better writers, some more knowledgeable about methodology, and others better conceptualizers. A skill that has often been noted is the ability to know one's strengths and capitalize on them and, conversely, to know one's weaknesses and minimize them. This ability may be enhanced by knowing when and how to avoid certain situations and to engage purposefully in other situations. Another tack is to work diligently on improving serious weaknesses. Skills critical to the domain may not be avoidable, and the student must determine those skills that cannot be compensated for by strengths in other areas.

An improved understanding of successful graduate students would have several implications. First, in contrast with the static evaluations of students that occur prior to matriculation (e.g., admissions tests) and at various crossroads of the graduate career (orals, defenses), it may be possible to assess student growth dynamically on a number of critical skills. Second, improved understanding of performance would increase the probability that assessment results would have instructional value. Third, understanding skill development and designing both instruction and

assessment based on this understanding can have profound effects in making graduate education more accessible. Testing that uses assessments of learned competencies only as predictors of later success, without instructional implications for improvement, does not benefit those who have not had the opportunities to develop certain skills.

Prediction and assessment. Willingham (1974) discussed the need for developing new and appropriate criteria of graduate school success in order to increase the validity of predictive measures. Criterion variables such as first year GPA or time-to-degree do not capture much of the critical information that determines graduate success. Hartnett and Willingham (1979) suggested that work samples or simulations be developed to serve as intermediate criteria of success in graduate school. In addition, Frederiksen (1986) has discussed how simulation testing could be used to develop more efficient and less costly assessment instruments with good construct validity. His approach will be described in more detail in Part II of this report.

Instruction. Traditional standardized assessment tests have had few implications for instruction in graduate school because the skills tested, although predictive of and contributing to graduate school learning, are not the same skills as those endorsed as desired outcomes of a graduate education. Few graduate faculty would consider it worthwhile to try to teach their students complex reasoning by teaching them to solve four-term verbal analogies, for example. However, if assessment is based on analysis of work samples that are related to the desired outcomes of graduate education, the possibility of a fruitful interaction between assessment and instruction is improved. In addition, it may be possible to profile students on relevant dimensions and then make instructional decisions based on the profile.

Finally, although some skills critical to graduate success are teachable, they may not be explicitly taught in the course of graduate education. This lack of explicitness may have differential consequences for students from various backgrounds, some of whom may not have had prior experiences important for the development of certain critical skills. If these weaknesses are to be overcome in a graduate program, explicit identification of these skills is necessary.

An important issue here is whether the socialization process, so dependent on modeling, is always the best form of training for all students. Specifically, modeling may be best when faculty and students are most alike in their backgrounds and outlook. Students with backgrounds and outlook different from faculty may not benefit as much from the modeling approach or may, in fact, be adversely affected. For a multitude of reasons, including the fact that all disciplines benefit from divergent thinking, it is important to adapt training to encourage potential scholars from all backgrounds. A

BEST COPY AVAILABLE

critical question is how graduate training can be modified to achieve this goal.

Guidance. From another point of view, dissemination of explicit information about what is expected from graduate students may help undergraduates make more informed decisions about attending graduate school, and about what types of programs to apply to. Such information would also help students evaluate the match between their interests and those of graduate programs. Furthermore, information obtained about the values and interests of successful students might provide a basis for the development of discipline-specific interest inventories that could aid students in making career decisions and aid faculty in coping with diversity in the student body.

Communication. A final aspect of graduate education that might be affected by this research is communication. A common language could be useful in describing individuals for job positions, promotional decisions, and comparative ratings of programs. Student evaluations can also be made against certain objective criteria.

Summary

In Part 1 of this report we presented a description of the educational process that occurs in graduate school, of the characteristics of successful graduate students, and of potential simulation tasks that would permit students to display these competencies. This description serves to enlarge our conceptualization of the behavioral domain to be assessed and of the ways in which it may be assessed. Graduate students need not only to reason well and to master large bodies of knowledge but also to add to this body of knowledge, to communicate knowledge to others, and to become active professionals. Opportunities to demonstrate these emerging competencies include proposal and report writing, and discussions and interactions with professors, other graduate students, and undergraduates. Potential simulation tasks that present such opportunities in a more standardized format were described.

However, in order for this kind of description to have an impact on the assessment process, a program in which these intuitive conceptualizations are grounded in empirical research is needed. For example, "communication skill" needs to be defined in such a way that various observers can agree on the degree that it is demonstrated in a particular instance. In Part II of this report we elaborate on how simulation testing can be used as the basis of such a research program that has as its goal the development of assessment materials that measure a wider variety of skills and serve a wider variety of assessment needs.

Part II: Simulation Testing and Test Development

Simulation testing has been used in many actual and experimental assessment programs and, less frequently, as part of a certification

process. Consequently, there exists a great deal of research concerning the psychometric properties of such tests. In the first two sections of Part II, we summarize some results from research on simulation testing in industrial assessment centers and in academia. This summary is, by no means, a comprehensive review of all relevant research. Such a review is well beyond the scope of the current project. Rather, it is designed to selectively illustrate some of the strengths and weaknesses of simulation testing as it has been used, and to document the variety of approaches to scoring simulations. We focus, first, on research in industry because that is where simulation testing has been most widely used and studied and, second, on research in academia because of its obvious relevance to the concerns of the current project. In the third section, we discuss the role simulation might play in the test development process.

Simulation Testing in Industry

The extensive body of research on simulation testing in industry provides an indication of the current state of the art in terms of psychometric considerations such as the reliability of exercises and scoring systems as well as their criterion and construct validity. However, before summarizing this research we briefly describe the assessment process typically used in industry, which relies heavily on clinical judgment.

The Assessment Process in Industry. The assessment center approach makes use of a team of trained assessors. Typically, several assessors observe participants in different exercises. The primary observer prepares an exercise report of behavioral observations relevant to the dimensions to be assessed on that exercise. Subsequently, all the assessors make independent ratings of the dimensions exhibited by the participant on that exercise based on the primary observer's behavioral report or on audiovisual recordings of the event. In the next stage of the assessment, the assessors independently integrate information from all the exercises as well as from other assessment procedures, such as paper-and-pencil tests, into overall dimension ratings. Subsequently, the assessors discuss their ratings on each dimension until a consensus is reached. In addition, after the complete record has been reviewed, an overall assessment rating is usually assigned to each participant by each assessor and inter-assessor differences in this rating are resolved through discussion. Thus the final assessment ratings are typically a product of clinical rather than statistical integration of many different sources of data, although statistical integration also occurs.

Psychometric Considerations. Given the complexity of the assessment center approach and its heavy reliance on clinical judgment, there are several psychometric issues that must be considered in evaluating this approach. These include the reliability and validity of overall performance ratings, the contribution of various assessment procedures to the overall performance ratings and to the prediction of job success, the comparative value of clinical

versus statistical prediction, and the reliability and validity of scores on assessment exercises as well as on behavioral dimensions. Evidence reviewed by Thornton and Byham (1982) concerning these various issues is briefly summarized below.

Overall assessment ratings (OARs) have been found to have good criterion validity. Thornton and Byham (1982) describe five well-designed longitudinal studies of the assessment center approach in which no feedback was provided to either the participants or to their supervisors. Assessors' ratings of participants' management potential correlated .30 to .50 with job success as measured by variables such as level of management attained or by salary progress over periods of 2.5 to 16 years. For example, in one study the correlations between assessors' prediction of management potential and actual attainment of middle management positions within six to eight years was .46 for both college and noncollege subjects. After 16 years the correlation was .33 for the college group and more than .40 for the noncollege group.

OARs are composites based on the results of a variety of assessment procedures, including paper-and-pencil tests, interviews, simulation exercises, and dimensional ratings. Therefore, the relative contribution of each of these sources of information to the OAR and to its criterion validity has been a question of interest. The results of studies of this issue indicate that the various assessment procedures contribute independently to both the OAR and to the prediction of success (see summary by Thornton & Byham, 1982). For example, Moses and Boehm (1975) reported that the correlations of management level were .44 with OAR, .32 (median) for various dimensional ratings from exercises, and .21 with the School and College Aptitude Test. Thus, information from some of the exercises as well as from the ability test contributed to the OAR and its prediction of the criterion.

A related question is whether data from the various assessment procedures should be combined clinically by human judges or mechanically through statistical procedures. There is a substantial body of research in psychological and medical diagnosis indicating that the ability of clinicians to synthesize multiple sources of information into a prediction of patients' status is inferior to statistical models (Wiggins, 1973). Although most assessment centers use clinical methods to make predictions, research contrasting the two methods in this context has not been extensive and has produced equivocal results. However, Wiggins' analysis clearly indicates that a statistical integration of data would be more appropriate.

Particular assessment techniques have been scrutinized to differing degrees and, consequently, information related to reliability and validity indices varies greatly. There have been numerous studies of the relationship of paper-and-pencil tests of ability and personality to performance on all kinds of jobs. Overall, the predictive validity for such tests typically ranges between .20 to .35, though negative correlations are sometimes found. On the other

hand, little research has been conducted concerning the reliability and validity of structured background interviews as they are currently used in assessment centers. Although the few studies that have been done suggest that they make a reliable, valid, and unique contribution to the assessment, not enough data are available to estimate the size of this contribution. Data on simulation exercises are sketchy. For example, both leaderless discussion groups and in-basket tasks are frequently used in assessment centers. Extensive research on leaderless discussion groups indicates that interrater reliability is typically .70 to .90 and predictive validity is between .30 and .50. However, there is little information about the reliability and validity of the in-basket test in the assessment center setting despite its widespread use.

In many assessment centers, behavioral dimension ratings across exercises are used. These ratings are usually obtained after the assessors have discussed the candidate's performance on all the exercises. Therefore, it is not surprising that the interrater reliability for dimension scores is typically .80 to.90. However, in one study (Schmitt, 1977) ratings were obtained both before and after discussion. Rater agreement was reasonably good before (median r = .67 for 17 dimensions) as well as after discussion (median r = .82).

Evidence relevant to the construct validity of the rated behavioral dimensions is mixed. On the one hand, a number of common factors have emerged in factor analytic studies of dimension ratings, including, for example, administrative skills, interpersonal skills, and an activity factor (energy, aggressiveness). On the other hand, the discriminant validity of many of these behavioral dimensions is poor. Correlations of ratings of the same dimensions across exercises tend to be lower than correlations of different dimensions on the same exercise.

There is evidence, however, that future attainment can be predicted by independent assessments of some of these skill dimensions. Performance ratings on the job have been found to correlate ($r > .30$) in at least two studies for each of the following: management skills, decision-making skills, communication, and interpersonal skills. Similarly, when job progress is the criterion, validity coefficients of at least .30 have been reported and replicated for decision-making skills, interpersonal skills, and other skills, such as initiative, independence, and self confidence.

Summary. The strengths of the assessment center approach lie in the documentation of skills and tasks important for job success, in the development of behavioral definitions of important skills so they can be reliably measured, and in the broad range of skills that are commonly assessed. The limitations of the assessment center approach thus far have been the insufficient attention paid to statistical integration of assessment information and to construct validity. However, the potential to explore these issues exists in the design of many assessment centers programs. The multitrait-multimethod matrix

BEST COPY AVAILABLE

design (Campbell & Fiske, 1959) of many of these studies offers an organizational framework for thinking about graduate school tasks and associated skills and for exploring empirically some aspects of the construct validity of the hypothesized competencies. The measurement of a number of hypothetically independent traits by a number of different methods provides an opportunity to examine the convergent and discriminant validity of the traits.

Simulation Tests in Academia

The relevance of simulation exercises for assessment in higher education has been explored in only a few areas. Three examples of research in academic settings are described in this section. Frederiksen and his colleagues have developed paper-and-pencil simulation tests of scientific thinking for use as criteria of creative problem solving. Simulated diagnostic interviews have been used for assessment in medical school and for physician credentialing. Crooks and her associates have investigated the use of in-basket tests in selection for admission to graduate schools of business. This work is described in more detail below.

Tests of Scientific Thinking. Frederiksen & Ward (1978) developed a set of Tests of Scientific Thinking (TST) to study problem-solving behavior in the context of activities normally undertaken by research psychologists. The tests were prepared for senior-level psychology students intending to pursue graduate training. The authors were particularly interested in developing intermediate criterion measures for use in research on creativity in problem solving and therefore provided subjects with a format intended to encourage production of unusual ideas. A departure was made from traditional test formats by presenting subjects with a set of realistic problems and asking them to write their own responses.

Since job analyses defining the domain of activities and situations involved in scientific research had not been undertaken, Frederiksen and Ward (1978) used an early description of scientists' work developed by Flanagan and his colleagues. Based on this analysis, four tests were developed simulating important tasks conducted by research scientists: Formulating Hypotheses, Evaluating Proposals, Solving Methodological Problems, and Measuring Constructs. A Formulating Hypotheses (FH) problem consisted of a brief description of a research study, a table or graph showing the results, and a statement summarizing the major finding. Examinees were instructed to generate hypotheses that might account for the finding and to indicate which of the hypotheses was the most likely explanation. Evaluating Proposals (EP) provided opportunities to critique research proposals. Examinees were given several research proposals ostensibly written by college seniors and were instructed to write critical comments for each student concerning the design, methodology, or theoretical position of the paper. Solving Methodological Problems (SMP) asked for suggested solutions to methodological problems in the planning of a research study, given brief statements of the problems.

In Measuring Constructs (MC), examinees were asked to suggest observable, measurable behaviors for operationalizing psychological constructs.

A scoring scheme was devised for each problem by classifying responses into categories, which were then ranked and assigned quality ratings by expert judges. Assignment of responses to categories yielded measures for average quality, number of responses, number of unusual responses, and number of unusual, high-quality responses.

There was considerable variation in the reliability estimates for scores within subjects and between raters. Measuring Constructs had the highest reliabilities for quality ratings and number of responses, with alphas equal to or exceeding .80. Solving Methodological Problems had the lowest reliabilities, with estimates below .63 for quality and about .73 for number of responses. When comparing estimates across tests, the number of responses was the most reliable measure while lowest reliabilities were found for the number of unusual and unusual, high-quality responses. These latter response categories, of course, also had the lowest frequencies of occurrence.

The predictive validity of TST was examined in a follow-up study of first-year graduate students in psychology (Frederiksen & Ward, 1978). GRE scores were the best predictors of first-year grades in graduate school, the traditional academic criterion. However, the number and unusualness of ideas were better predictors of self-reported involvement in professional behaviors such as attending professional meetings, publishing, and engaging in collaborative research. Thus, the TST have some validity in predicting engagement in future professional activities.

The construct validity of the TST has also been studied. Using both correlational and factor analytic techniques, the relationships of scores on these tests to performance on the GRE verbal, quantitative and Advanced Psychology tests were explored (Frederiksen & Ward, 1978; Ward, Frederiksen, & Carlson, 1980). As a substantial amount of the reliable variance in the TST was not associated with performance on any GRE test, it would appear that the TST measured somewhat different abilities.

The construct validity of the Formulating Hypotheses test was further explored by Ward, Frederiksen and Carlson (1980). A set of 12 cognitive and personality variables hypothesized to contribute to the solution of these problems was identified. Moderate correlations were found between scores on a free-response version of FH and the cognitive and personality measures. However, the GRE aptitude and achievement tests showed stronger relationships with verbal, reasoning, and cognitive flexibility than did the FH test; nonetheless, FH, in the free-response format, was distinct in demonstrating a relationship to fluency factors. Moreover, correlations between FH tests in free-response and multiple-choice formats were low, suggesting that different processes were used for

BEST COPY AVAILABLE

administered in a free-response format, appears to measure some aspects of creative thinking that traditional tests do not.

Simulated Interviews. Among the best examples of this simulation testing approach are the Patient Management Problems (PMP) used by medical schools to evaluate clinical performance of students and by medical boards to credential physicians. Intended to be realistic work samples of clinical situations, these problems present some information about a patient that is followed by cycles of asking further questions of the patient, or ordering laboratory tests. On the basis of feedback regarding these inquiries, hypotheses about possible diagnoses are to be suggested by examinees. Typically, PMPs employ a multiple-choice format and are scored by some variation of a number-correct formula. However, examinees may be assessed in terms of their efficiency and accuracy of diagnosis, and errors in patient management may be noted (Elstein, Shulman, & Sprafka, 1978).

Braun, Carlson, and Bejar (1987), in summarizing much of this research, seriously question the validity of these instruments. Sampling problems, the lack of consistency across studies, the lack of refined psychometrics to analyze this type of data, and the lack of good criterion measures all contribute a litany of criticisms, not the least of which is poor generalizability across different patient-management problems. While the face validity of PMPs is compelling, serious questions abound concerning criterion and construct validity.

In-basket Performance. An in-basket simulation, employing a free-response format, was used in a study of entering MBA students to determine if this approach would differentially predict grades and faculty ratings during graduate school (Crooks, 1971). Students were presented with an in-basket of administrative tasks along with biographical and personality questionnaires. Their performance on the in-basket test was scored on a number of dimensions, including, for example, taking action, problem analyzing, delegating, and amount of work accomplished. Achievement measures were obtained from student records at the end of their first year of graduate study. In addition, faculty were asked to rate students. In-basket performance did not improve the prediction of first year-grades above traditional measures. However, when faculty ratings of students were the criterion, the multiple correlation increased from .13 to .24 when in-basket performance was added to the prediction equation. Cluster analyses suggested that in-basket scores were providing information about a student's skills distinct from traditional predictors of student academic performance.

Summary. Research within higher education settings has provided some evidence that the information obtained about cognitive performance through the use of simulations, particularly when a free-response format is used, is different from that obtained through traditional paper-and-pencil procedures. Further, there is some evidence that such simulation measures have the potential to predict performance in real-world professional activities. Thus, the

advantages of simulation testing noted earlier in industrial settings apply to academic settings as well. Simulations offer an opportunity to document empirically a broader range of skills and task performances that are important for success in real-world activities and situations.

Nevertheless, serious problems still exist with such assessment approaches. The nature of these problems and the research directions that will contribute to the their resolution are discussed below.

Simulation Testing and the Development of Construct Valid Tests

It is clear that simulation testing cannot easily be adapted to large-scale standardized testing at present. Among the major obstacles that would have to be overcome are problems of cost and efficiency, of psychometric immaturity, and of inadequate attention to the construct validity of the tasks. With respect to cost and efficiency, most simulation tests are very time-consuming in terms of both examinee and assessor time. In terms of psychometric issues, the problems are legion because most psychometric models assume item independence. One example of such a problem is that the reliability of Patient Management Problems, in which responses to items are sequentially dependent, cannot be appropriately estimated by conventional methods of measuring reliability that assume independence of items. Furthermore, problems of efficiency and psychometric defensibility can interact. For example, if problems are too time-consuming, it will restrict the sample of problems that can be administered and thereby reduce both reliability and content validity. A related issue is the trade-off between realism and control, between the need for sufficient complexity to engage the processes relevant to real-world tasks and the need for sufficient standardization to sustain the construct validity and comparability of scores across examinees. Finally, the intuitive appeal of simulation tests has often led to a reliance on face and content validity as a substitute for construct validity.

Nevertheless, the promise of simulation tests is strengthened by the advances that are currently being made in technological and theoretical areas. Powerful, intelligent computers as well as other delivery systems can already provide a cost-efficient method of presenting simulation problems and of recording and scoring some types of student responses. As the computer technology for processing natural language improves, it may become feasible to automatically assess the quality of free responses. For example, Carlson and Ward (1987) have explored the potential for computer administration of Formulating Hypotheses items. They concluded that a prototype system for test delivery and for scoring of open-ended, sentence-level responses can be achieved with currently available tools and techniques.

The problem of construct validity, however, can be seen as a central one in that it will be difficult to design efficient tests or

reliable scoring systems unless critical aspects of performance associated with expertise are identified. An example of how attention to the construct validity of a simulation task can foster new directions for test development is provided by Braun, Carlson, and Bejar (1987). Their review of research on PMPs led them to conclude that the data-gathering components on such tasks (i.e., seeking information) contribute little to the discrimination of level of expertise. Instead, decision-making processes appear to be more important and Braun and his colleagues offer suggestions as to how such skills could be assessed. A longitudinal data base including measures of performance on simulation tasks would facilitate research on skill development, on the assessment of developing expertise, and on the prediction of success. If the constructs measured in the assessment instrument are the same as those important to successful performance on real-world tasks, strengths and weaknesses detected by the assessment will have direct implications for performance in the graduate context. Furthermore, if the construct and criterion validity of the simulation exercises is established, they could be used as intermediate criteria for other tests.

Frederiksen (1986) has advocated the use of simulation testing in the test development process. His programmatic approach to the development of assessment instruments can be seen as having two phases. The outcome of the first phase is the development of construct valid simulation tasks that can serve as criteria in the development of other tests; the outcome of the second phase is the development of more efficient, less costly operational tests.

The first phase involves three steps. The first step is to construct a theory of criterion performance or a model of the skills involved in successful performance. This model may be based on a review of relevant research, if any exists, or on task or job analyses. The second step in Frederiksen's program is to develop assessment tasks that parallel as closely as possible the real-life performance to be assessed. The third step involves a test of the model and the establishment of the construct validity of the assessment tasks. Testing the model and exploring the construct validity of the simulation tasks involves collecting work samples, developing scoring systems that operationalize the processes and skills hypothesized in the first step, and collecting other data that would help explicate the meaning of the scores on the work sample. A variety of means can be used to establish the construct validity of the simulation tasks. These include investigations of the relationships among the skills hypothesized and other indicators of target constructs, experimental demonstrations that similar processes underlie successful performance on simulation and real life-tasks, and evidence that performance on the tasks improves with training or varies with level of expertise. For our purposes, in addition to developing intermediate criteria for use in validating operational tests, these steps can also function as a research program that can lead to a theory of successful performance in graduate school.

The second phase of this program is concerned with the development and construct validation of more efficient operational tests. After such tests are developed, they need to be validated by demonstrating their construct similarity to the simulation tests. In addition, construct similarity for trained and untrained groups should be documented because the processes involved in completing a task may vary with level of expertise.

In practice this approach has been used to assess fairly discreet skills in isolation, for example, creative problem-solving or clinical diagnosis, for which a reasonable theory of performance already exists. However, one goal of a research program concerning graduate school performance should be to adapt this approach to investigate more complex skills that occur in conjunction on more complex tasks. Higher-level skills that are used to coordinate more basic skills may not be evident on simplified versions of tasks. A major challenge here will be in developing methods of analyzing simulations that can reliably differentiate performance along a continuum of expertise on a number of different dimensions. Various approaches to the analysis of simulation exercises are described below.

Systems that have been used for analyzing simulation exercises or that might be adapted for such purposes range from very global to very analytic. The overall assessment ratings used in assessment centers or holistic systems for scoring essay tests (Cooper, 1984) are examples of global scoring systems. These systems result in a single, overall score summarizing performance on a task or on a series of tasks as a whole. The category scoring system used to assess creative problem solving (Frederiksen & Ward, 1978) or the behavioral dimension ratings used in assessment centers represent an intermediate type of system. In such systems, a few selected performance characteristics are rated or scored separately. An example of a highly analytic approach is that used by Voss and his associates (Voss, Greene, Post, & Penner, 1983) in their investigations of problem solving in the social sciences. They conducted highly detailed analyses of protocols in which nearly every idea expressed by the problem solver was classified in terms of its function in a goal structure and a reasoning structure. This type of analysis enabled Voss and his colleagues to identify a number of differences in the way that experts and novices approached social science problems, including the generality of the problem representation and the amount of argumentation offered to support a solution.

Obviously, these approaches differ in their efficiency as well as their utility. Global systems are most efficient in terms of scorer time and may be the most reliable, but they provide little information about why performance is good or bad. Therefore, such systems may be useful in assessment programs that are primarily concerned with prediction. Highly analytic systems are, at present, inefficient and not very reliable. Their potential usefulness is as research tools that can be used to identify critical aspects of performance so that more efficient tests and scoring systems can be developed.

Concluding Comments

In our discussions with graduate faculty, we tried to identify critical skills associated with scholarly and professional competence that are not currently assessed by graduate admissions tests. Moreover, we attempted to develop recommendations about a research approach that would foster the empirical definition of these skills in such a manner that feasible methods of assessing them might eventually be developed. The idea of simulation testing emerged as a central theme in our discussions because of its usefulness as an organizing framework, its promise as a research approach, and its potential as an alternative approach to assessment that deserves serious consideration. Although it is clear that many obstacles need to be overcome before simulation testing can be used in large-scale testing programs, there are many immediate benefits that would derive from a program of research based on simulation testing. First and foremost, it would help to clarify the nature of variables that contribute to success and achievement in graduate school. Second, simulation tests might be used as criteria in evaluating current tests. Third, such a research program would stimulate the development of new types of assessment instruments that might serve a wider variety of purposes. These instruments might be more efficient or economical versions of simulation exercises that focus on critical aspects of task performance, or they might be new types of instruments, such as discipline-specific interest inventories that could be used for guidance rather than admissions purposes. Moreover, a standardized assessment program based on simulation testing might become more attractive if such tests provide educationally relevant information about the development of students over the course of their education. Finally, such a program should make an important contribution to communication among undergraduate programs, potential graduate students, graduate students, and graduate programs. The definition of the competencies that graduate students are expected to develop and the tasks they are expected to accomplish will make it easier for undergraduate programs to prepare students appropriately, will better inform potential graduate students about the nature of graduate education, and will permit both graduate students and graduate faculty to better assess both student progress and program effectiveness.

BEST COPY AVAILABLE

References

Baird, L. L. (1979). Development of an inventory of documented accomplishments for graduate admissions. (Graduate Record Examinations Board Research Report GREB No.77-3.) Princeton, NJ: Educational Testing Service.

Baird, L. L. (1985). Field trial of a user-oriented adaptation of the inventory of documented accomplishments as a tool in graduate admissions. (Graduate Record Examinations Board Research Report GREB No.81-1r, ETS Research Report 85-13.) Princeton, NJ: Educational Testing Service.

Baird, L. L., & Knapp, J. E. (1981). The inventory of documented accomplishments for graduate admissions: Results of a field trial study of its reliability, short-term correlates, and evaluation. (Graduate Record Examinations Board Research Report GREB No. 78-3, ETS Research Report 81-18.) Princeton, NJ: Educational Testing Service.

Bartholomae, D. (1986). Inventing the university. Journal of Basic Writing, 5, 4-23.

Braun, H., Carlson, S., & Bejar, I. I. (1987). Psychometric foundations of testing based on patient management problems. Unpublished manuscript.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 56, 81-105.

Carlson, S. B., & Ward, W. C. (1987). A new look at Formulating Hypotheses items. (Graduate Record Examinations Board Research Report GREB No. 85-14, in preparation.) Princeton, NJ: Educational Testing Service.

Chi, M. T. H., Feltovich, P. J. & Glaser, R. (1981). Categorization and representation of physics knowledge by experts and novices. Cognitive Science, 5, 121-152.

Chi, M. T. H.; Glaser, R.; & Farr, M. J. (1988). The Nature of Expertise. Hillsdale, NJ: Erlbaum.

Conrad, L. (September 1976). Aptitude test restructuring research: Findings concerning projected development of an abstract reasoning measure. (Unpublished technical report for the Graduate Record Examinations Board.) Princeton, NJ: Educational Testing Service.

Cooper, P. L. (1984). The assessment of writing ability; a review of the research. (Graduate Record Examinations Board Research Report GREB No. 82-15R.)

Crooks, L. A. (1971). The in-basket study. (ATGSB Brief No. 4.) Princeton, NJ: Educational Testing Service.

Crooks, L. A., & Campbell, J. T. (April 1974). Career progress of MBAs: An exploratory study six years after graduation. (ETS Progress Report No. 74-8.) Princeton, NJ: Educational Testing Service.

Crooks, L. A.; Campbell, J. T.; & Rock, D. (1979). Predicting career progress of graduate students in management. (ETS Research Report No. 79-15.) Princeton, NJ: Educational Testing Service.

Dawes, R. M. (1971). A case study of graduate admissions: Applications of three principles of human decision making. American Psychologist, 26, 180-188.

Donlon, T. F.; Reilly, R. R.; & McKee, J. D. (1978). Development of a test of global vs. articulated thinking: The Figure Location Test. (Graduate Record Examinations Board Report 74-9p). Princeton, NJ: Educational Testing Service.

Elstein, A. S.; Shulman, L. S.; & Sprafka, S. A. (1978). Medical problem solving: An analysis of clinical reasoning. Cambridge, MA: Harvard University Press.

Flanagan, J. C. (1954). The critical incident technique. Psychological Bulletin, 51, 327-358.

Frederiksen, N. (1984). The real test bias: Influences of testing on teaching and learning. American Psychologist, 39, 193-202.

Frederiksen, N. (1986). Construct validity and construct similarity: Methods for use in test development and test validation. Multivariate Behavioral Research, 21, 3-28.

Frederiksen, N., & Ward, W. C. (1978). Measures for the study of creativity in scientific problem-solving. Applied Psychological Measurement, 2, 1-24.

Gentner, D., & Stevens, A. L. (1983). Mental models. Hillsdale, NJ: Erlbaum.

Glaser, R. (1981). The future of testing: A research agenda for cognitive psychology and psychometrics. American Psychologist, 36, 923-936.

Grice, P. (1975). Logic and conversation. In P. Cole and J. L. Morgan (Eds.), Syntax and semantics (Vol. 3). New York: Academic Press.

Hartnett, R. T., & Willingham, W. W. (1979). The criterion problem: What measure of success in Graduate education? (Graduate Record Examinations Board Report 77-4R.) Princeton, NJ: Educational Testing Service.

Hilton, T. L.; Kendall, L. M.; & Sprecher, T. B. (1970). Performance criteria in graduate business study, Parts 1 and 2: Development of rating scales, background data form and the pilot study. (ETS Research Bulletin No. 70-3.) Princeton, NJ: Educational Testing Service.

Johnson-Laird, P. N. (1983). Mental models. Cambridge, MA: Harvard University Press.

Larkin, J.; McDermott, D.; Simon, D.; & Simon, H. A. (1980). Expert and novice performance in solving physics problems. Science, 208, 1335-1342.

Lesgold, A. M. (1984). Acquiring expertise. In J. R. Anderson and S. M. Kosslyn (Eds.), Tutorials in learning and memory: Essays in honor of Gordon Bower. San Francisco: W. H. Freeman.

McCloskey, M. (1983). Naive theories of motion. In D. Gentner and A. L. Stevens (Eds.), Mental models. Hillsdale, NJ: Erlbaum.

Moses, J. L., & Boehm, V. R. (1975). Relationship of assessment center performance to management progress of women. Journal of Applied Psychology, 60, 527-529.

Powers, D. E., & Enright, M. K. (1987). Analytical reasoning skills involved in graduate study: Perceptions of faculty in six fields. Journal of Higher Education, 58, 658-682.

Reilly, R. R. (1976). Factors involved in graduate student performance. American Education Research Journal, 13, 125-138.

Schmitt, N. (1977). Interrater agreement in dimensionality and combination of assessment center judgments. Journal of Applied Psychology, 62, 171-176.

Schoenfeld, A. H. (1985). Mathematical problem solving. Orlando: Academic Press.

Sternberg, R. J. (1985). Beyond IQ. New York: Cambridge University Press.

Thornton, G. C., III, & Byham, W. C. (1982). Assessment centers and managerial performance. New York: Academic Press.

BEST COPY AVAILABLE

Tucker, C. (1985). Delineation of reasoning processes important to the construct validity of the analytical test. (Unpublished report for the Graduate Record Examinations Board.) Princeton, NJ: Educational Testing Service.

Voss, J. F.; Greene, T. R., Post, T. A.; & Penner, B. C. (1983). Problem-solving skill in the social sciences. In The Psychology of Learning and Motivation (Vol. 17), pp. 165-213. New York: Academic Press.

Ward, W. C.; Frederiksen, N.; & Carlson, S. B. (1980). Construct validity of free-response and machine-scorable forms of a test. Journal of Educational Measurement, 17, 11-29.

Wiggins, J. S. (1973). Personality and prediction: Principles of personality assessment. Reading, MA: Addison-Wesley.

Willingham, W. W. (1974). Predicting success in graduate education. Science, 183, 273-278.

Appendix
Faculty Consultants

Dr. Nancy S. Barrett
Department of Economics
American University

Dr. Gordon Bower
Department of Psycholology
Stanford University

Dr. Benjamin DeMott
Department of English
Amherst College

Dr. Lloyd Ferguson
Department of Chemistry
California State University,
Los Angeles

Dr. Eric Foner
Department of History
Columbia University

Dr. Robert Glaser
LRDC
University of Pittsburgh

Dr. Edmund Gordon
ISTS
Yale University

Dr. Earl Hunt
Department of Psychology
University of Washington

Dr. George A. Miller
Department of Psychology
Princeton University

Dr. Gilbert Rozman
Department of Sociology
Princeton University

Dr. Robert Sternberg
Department of Psychology
Yale University

Dr. Patrick Suppes
Department of Philosophy
Stanford University

Dr. William Thurston
Department of Mathematics
Princeton University

Dr. Evelyn Witkin
Waksman Institute of
Microbiology
Rutgers University

Dr. Sheldon Wolin
Department of Politics
Princeton University

54020-02951 • Y50P1.5 • 297984